The persistence of depression score

Spijker, J.; de Graaf, R.; Ormel, J.; Nolen, W. A.; Grobbee, D. E.; Burger, H.

Published in:
Acta Psychiatrica Scandinavica

DOI:
10.1111/j.1600-0447.2006.00821.x

Acta Psychiatr Scand 2006: 114: 411–416
All rights reserved
DOI: 10.1111/j.1600-0447.2006.00821.x

ACTA PSYCHIATRICA SCANDINAVICA

The persistence of depression score

Spijker J, de Graaf R, Ormel J, Nolen WA, Grobbee DE, Burger H. The persistence of depression score.

Objective: To construct a score that allows prediction of major depressive episode (MDE) persistence in individuals with MDE using determinants of persistence identified in previous research.
Method: Data were derived from 250 subjects from the general population with new MDE according to DSM-III-R. These subjects were recruited from the Netherlands Mental Health Survey and Incidence Study (NEMESIS), using the Composite International Diagnostic Interview. Determinants for persistence were transformed into a practical risk score using proportional hazards models and bootstrapping techniques.
Results: The risk of MDE persistence after 12 months was 23%. The score comprised measures of physical illness, social support, depression severity and recurrency, and duration of previous episodes. With increasing categories of these measures, the predicted risks increased from 7 to 40%.
Conclusion: We constructed the Persistence of Depression Score (PDS) showing reasonable performance. The PDS could be of importance in clinical practice to support treatment decisions.

J. Spijker[1,2], R. de Graaf[1], J. Ormel[3], W. A. Nolen[3], D. E. Grobbee[4], H. Burger[3]

[1]Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands, [2]De Gelderse Roos, Institute for Mental Health Care, Wolfheze, The Netherlands, [3]Department of Psychiatry, University Medical Centre, Groningen, The Netherlands and [4]Julius Centre for Health Sciences and Primary Care, University Medical Centre, Utrecht, The Netherlands

Key words: depression; prognosis; risk assessment; general population

Jan Spijker, Netherlands Institute of Mental Health and Addiction (Trimbos Institute), PO Box 725 3500 AS Utrecht, The Netherlands.
E-mail: jspijker@trimbos.nl

Accepted for publication March 22, 2006

Significant outcomes

- The risk of persistence of major depressive episodes after 12 months is 23%.
- The persistence of depression score (PDS) is a clinical risk score to predict persistence of depression with reasonable psychometric performance.
- The PDS can be applied with relative low burden to patient and budget.

Limitations

- Potential important genetic and biological predictors of persistence of depression were not assessed.
- The external validity and usefulness of the PDS needs to be studied further.

Introduction

Persistence of a major depressive episode (MDE) is a common, and serious, clinical problem (1). If its risk can be estimated accurately in single patients, treatment may be individualized in the sense that those with the highest risk receive most intense treatment. However, setting a prognosis in individual cases is difficult. The prognostic skills of clinicians are still underdeveloped and lack an empirical basis (2). Although many studies of depressed patients identified predictors of depression persistence, the analyses and presentation of the results do not allow to predict the risk in an individual patient (3–9).

In a recent paper based on results from a general population survey [Netherlands Mental Health Survey and Incidence Study (NEMESIS)], we identified the following independent risk factors for persistence of MDE: higher severity of the index episode, longer duration of previous episodes, (chronic) physical illness and lack of social support; a recurrent episode predicted earlier recovery (10).

Aims of the study

In the current paper, we will present a model how to combine these determinants into a score that enables a clinician to estimate the risk of persistence of depression in individual subjects with MDE.

Material and methods

NEMESIS is a prospective psychiatric epidemiologic survey in the Dutch adult general population (aged 18–64 years) with three waves in 1996 (T_0), 1997 (T_1) and 1999 (T_2). Methods are described in detail elsewhere (10–14). Briefly, NEMESIS is based on a multistage, stratified, random sampling procedure. In the first wave, sufficient data were gathered on 7076 persons, with a response rate of 69.7%. At T_1, 1458 respondents (20.6%) were lost to attrition and at T_2, a further 822 (14.6%) were lost leaving 4796 respondents that were interviewed at all three waves. Psychopathology did not have a strong impact on attrition: at T_1 12-months agoraphobia (odds ratio (OR), 1.96) and social phobia (OR, 1.37), and at T_2 12-months major depression (OR, 1.37), dysthymia (OR, 1.80) and alcohol dependence (OR, 1.83), adjusted for demographic factors, were associated with attrition (15).

Diagnoses of psychiatric disorders according to DSM-III-R (16) were based on the Composite International Diagnostic Interview (CIDI), Version 1.1 (computerised version (17)). The CIDI is a structured interview developed by the World Health Organization (18), which has been found to have acceptable interrater reliability and test–retest reliability for most diagnoses, including major depression (19). For our study cohort, we included respondents with newly originated MDE (first or recurrent cases) between T_1 and T_2, that is those with a diagnosis of 2-year prevalence of major depression at T_2 but no 1-month prevalence of major depression diagnosis at T_1 ($n = 250$). Individuals diagnosed with bipolar disorder or a primary psychotic disorder were excluded. Duration of MDE was assessed with the Life Chart Interview (20). A median duration of MDE of 3 months (95% CI 2.2–3.8 months) was found; for nearly 20% of the depressed subjects duration of the episode exceeded 24 months (13). A broad range of potential determinants of persistence were examined: age and gender, educational attainment, cohabitation status, negative youth experiences, neuroticism, previous psychiatric illnesses, physical illness, life events and ongoing difficulties, social support, and clinical characteristics as severity of depression, the index episode being a first or recurrent episode, duration of previous episodes and comorbidity with other DSM-III-R axis I disorders [see Ref. 10 (10)].

Analyses

Cox proportional hazards models were used to estimate the associations between the predictors and recovery from MDE as published (10). For the present study, the same final Cox proportional hazards model was fitted on the same data. The aim of the analysis was to calculate the cumulative 12-month risk of not having recovered, i.e. the risk of depression persistence $S(12\text{ months})$ for each patient. This appeared not straightforward as the Cox regression procedure yields the actual survival estimates [$S(t)$] only. These estimates represent the predicted risks according to the model of depression persistence for each patient with his given follow-up time and predictor values. They are defined as:

$$S(t) = S_0(t)^{\exp(\mathrm{LP})}$$

where the linear predictor (LP) is $\beta_1{*}X_1 + \beta_2{*}X_2 + \cdots$, with the X denoting the predictor values and β the regression coefficients. The baseline survival function $S_0(t)$ is the time-dependent cumulative risk of persistence of depression for a person with none of the predictors present, i.e. the LP being zero and thus $S_0(\mathrm{t}) = [\mathrm{S(t)}]$. The baseline survival function [$S_0(t)$] can be calculated by remoulding the above formula as follows:

$$S_0(t) = S(t)^{1/\exp(\mathrm{LP})}$$

This calculation allowed us to read the cumulative 12-month baseline risk from the database from those patients who actually had 12 months of follow-up [$S_0(12\text{ months})$]. In our study, this value appeared to be 0.2029. The final step was to calculate the 12 month risk for all patients using the $S_0(12\text{ months})$ and the LP, the latter thus representing the individual part of the risk. Hence,

$$\begin{aligned} S(12\mathrm{months}) &= S_0(12\mathrm{months})^{\exp(\mathrm{LP})} \\ &= S_0(0.2029)^{\exp(\mathrm{LP})}. \end{aligned}$$

The 12-month time span was arbitrarily chosen, not only primarily on clinical grounds but also because at the time of follow-up the number of patients at risk (of recovery) was still sufficiently large. To evaluate the calibration of the model, i.e. to assess the extend to which the model predictions are in agreement with the observed probabilities,

we calculated the Kaplan–Meier estimate of the 12-month risk of depression persistence for each decile of predicted risk and compared them using a scatter diagram (21).

As a next step, the discriminatory power of the model was quantified. Because the outcomes of the censored patients are unknown, the construction of a receiver operating characteristic (ROC)-curve for the evaluation of discriminatory power is impossible. However, the concordance statistic can still be calculated (22). This statistic is numerically, and as regards interpretation, equal to the area under the ROC-curve. It is the probability that for a random pair of patients the one who has the event first has the highest predicted probability. The concordance statistic (or area under the ROC-curve) is an overall measure of discriminatory power, with a value 0.5 indicating no discrimination, and a value 1.0 indicating perfect discrimination between those with and without the study outcome, i.e. depression persistence (23). Both the regression coefficients, and therefore also the hazard ratios with their 95% confidence intervals, as well as the concordance-statistic, were adjusted for overfitting or overoptimism using bootstrapping techniques (24). To this end, 100 random bootstrap samples with replacement were drawn from the dataset with complete data on all predictors ($n = 250$). The model's predictive performance after bootstrapping is the performance that can be expected when the model is applied to future similar populations (25).

To construct a practically applicable persistence of depression score (PDS), each coefficient from the model was transformed to a round number of points. As the coefficients reflect the relative weight of each variable in the prediction, they were transformed to a number of points in a uniform way, i.e. each coefficient was divided by the coefficient closest to zero, i.e. −0.107. The number of points was subsequently rounded to the nearest integer. The total score for each individual patient was determined by assigning the points for each variable present, and adding them up.

The predicted probability of persistence of depression at 12 months follow-up was presented according to four broad categories of the risk score for reasons of statistical stability and practical applicability. The categories were arbitrarily chosen with a view to reasonable size as well as clinical sensibility. Next, the score was transformed to dichotomous 'prognostic tests' allowing each patient to be classified as high or low risk of depression persistence. Sensitivity, specificity as well as the positive and negative predictive value of these tests were calculated for the same cut-offs of the PDS as those used to delineate the score categories. The data were analysed using SPSS 12.0 and S-PLUS 2000 (SPSS Inc., Chicago, IL, USA).

Results

Characteristics of the study population are displayed in Table 1.

Follow-up time ranged from 0.5 to 24 months and 187 subjects of the total population ($n = 250$) recovered. The final proportional hazards regression model appeared reasonably calibrated as the predicted and observed probabilities were similar over the whole range (Fig. 1). The shrinkage factor

Table 1. Sociodemographic and clinical characteristics of a cohort ($n = 250$) with newly originated major depressive episodes (first or recurrent) in the general population

Variable	%
Sociodemographic variables	
Gender (female)	66.8
Age (years)	
18–24	6.4
25–34	36.4
35–44	26.8
45–44	21.1
55–64	9.2
Education	
Low	3.6
Medium	37.6
High	31.2
University	27.6
Living with partner (yes)	63.6
Paid employment (yes)	70.0
Clinical variables	
Severe depression	30.4
Recurrent depression	43.2
Comorbid dysthymia	10.0
Comorbid anxiety disorder	34.0
Comorbid substance abuse/dependence	10.4

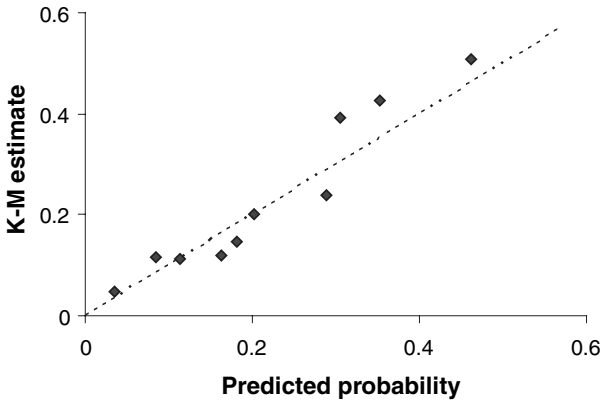

Fig. 1. Calibration plot of the Cox proportional hazards model for the prediction of depression persistence at 12 months of follow-up. The dotted line represents the line of identity, i.e. perfect model calibration.

Table 2. Multivariable predictors of recovery from depression

Predictor	Coefficient	Hazard ratio (95% confidence interval)	Contribution to risk score
Physical illness	−0.319	0.73 (0.54–0.97)	3
Medium social support	−0.107	0.90 (0.64–1.27)	1
Low social support	−0.420	0.66 (0.46–0.95)	4
Severe depression	−0.314	0.73 (0.53–1.01)	3
Recurrent depression	0.392	1.48 (1.10–1.99)	−4
Long duration previous episodes	−0.426	0.65 (0.48–0.89)	4

Total risk score = physical illness*3 + medium social support + low social support*4 + severe depression*3 − recurrent depression*4 + long duration previous episodes*4

for the coefficients that was obtained from the bootstrap process appeared 0.91. Models results are presented after shrinkage. Coefficients from the model as well as the hazard ratios as measures of relative risk are displayed in Table 2, together with the risk points per predictor. The total risk score was calculated using the formula at the bottom of the table. For instance, a subject with a severe and recurrent MDE, with a comorbid physical illness and low social support, has a score of $+3 - 4 + 3 + 4 = 6$ points.

Table 3 shows the relationship between categories of the score, the observed risk and the predicted risk of MDE persistence after 1 year. The mean risk was 23% and the predicted risks increased from 7 to 40% with increasing score categories and were generally well in agreement with the observed risk. From this table it can also be seen that the subject introduced above has a risk of 29% of persistence. The overall discriminatory power of the score was fair with a concordance statistic of 0.68. For specific cut-offs, the sensitivity, specificity and predictive values are presented in Table 4. If, for instance, a cut-off ≥5 is chosen as the threshold for more intense treatment, 69% (sensitivity) of those who would still suffer depression after 1 year will have received this treatment as their prognostic test was positive, i.e. high risk of persistence. Yet, 12% (1−NPV) of those who did not have the more intense treatment because their test was negative will still have depression.

Table 3. Risk of 12-month depression persistence according to score categories

Score	*n* (%)	Recovered (%)	Censored (%)	Observed risk (Kaplan–Meier estimate) (%)	Predicted probability (%)
<2	66 (26)	60 (32)	6 (10)	8	7
2–4	75 (30)	61 (33)	14 (22)	17	18
5–7	60 (24)	40 (22)	20 (32)	29	29
≥8	49 (20)	26 (14)	23 (37)	46	40
Total	250 (100)	187 (100)	63 (100)	23	22

Table 4. Prognostic test characteristics for 12 month depression persistence

Cut-off score	*n* (%)	Sensitivity (%)	Specificity (%)	PPV (%)	1−NPV (%)
≥2	184 (74)	93	32	27	7
≥5	109 (44)	69	63	34	12
≥8	49 (20)	36	85	40	17

PPV, positive predictive value; NPV, negative predictive value.

Discussion

In this paper, we extended our previous analysis of risk factors for persistence of MDE in the general population with the construction of the persistence of depression score (PDS). Especially in general practice, with a high prevalence of depressive disorder, such a risk score could be of clear importance in daily practice to support treatment decisions (26, 27). For instance, in a patient with a recently developed MDE and a low PDS score, a policy of watchful waiting might be adopted, as the natural course seems to be favourable. In a patient with a high PDS score, however, immediate and possibly more aggressive treatment is indicated (28).

A strength of our study is that the elements of the PDS can be assessed with relative low burden to patient and budget. However, the discriminatory power of the PDS is modest with a concordance statistic of 0.68, in particular when compared with concordance statistic or, equivalently, areas under the curve of the ROC curve obtained in diagnostic studies. But, it must be kept in mind that by nature of the close temporal relation between predictors and outcome, measures of discrimination generally achieve higher values in the diagnostic than the prognostic setting.

The concordance statistic is an overall measure of model performance, i.e. independent of a certain cut-off value for the predicted risk. In clinical practice, however, a sensible cut-off value must be chosen so that after the test, the probability of no remission will be very high or low that it helps to decide whether more intensive treatment (e.g. an antidepressant in combination with psychotherapy) is indicated. It must be kept in mind that false negative patients in a prognostic test, i.e. patients

that incorrectly were withheld the appropriate treatment will remain in treatment, so that treatment adjustments can be made, albeit with some time lag.

In NEMESIS we included individuals from the general population with depressive episodes, thereby avoiding referral filter bias. Some of these individuals received treatment, in primary care or in specialized mental health care (13). Depressed individuals from the general population differ from selected depressed in-patients or out-patients (12) and this could limit the utilization of the PDS in specific treatment settings. However, the factors we found to be associated with persistence, i.e. (chronic) physical illness, lack of social support and illness-related factors such as severity of the index episode and prior episodes and their duration, were very much in accordance with the literature and this applies certainly for the illness-related factors such as severity of the index episode and duration of prior episodes (3–9). Furthermore, no difference in duration of MDE was found in differential levels of care (13).

A further strength of the study is that we used bootstrapping techniques to correct for the inevitable problem of overoptimism of model performance in the derivation data, and consequently poorer predictive accuracy in new patients. Bootstrapping techniques have shown superiority over other approaches to address these problems such as split-sample or cross-validation methods (29). However, demonstrating the external validity and usefulness of this prediction score in the treatment of depressed patients in future studies remains indicated.

A limitation is that measurement error in the assessment of the duration of MDE by interview may have occurred and that as a result, some short recurrent episodes may have been gone undetected. Further, the predictors were analysed as dichotomous variables which inevitably has caused some loss of detail. However, a relatively crude assessment reflects clinical practice and therefore adds in our view to the generalizability of the results. A further limitation may be that NEMESIS did not include potentially important genetic and other biological predictors of persistence of MDE. Future research may yield biological predictors of persistence, which may improve our ability to forecast the course of depression.

In conclusion, we were able to construct a clinical risk score for the prediction of persistence of MDE with reasonable performance. The PDS may be of value to clinical practice in providing a rational basis for treatment decisions.

Acknowledgement

NEMESIS is supported by the Netherlands Ministry of Health, Welfare and Sport (VWS), the Medical Sciences Department of the Netherlands Organisation for Scientific Research (NWO), and the National Institute for Public Health and Environment (RIVM).

References

1. Judd LL The clinical course of unipolar major depressive disorders. (Commentary). Arch Gen Psychiatry 1997;**54**:989–991.
2. van den Brink RHS, Ormel J, Tiemens B et al. Accuracy of general practioners' prognosis of 1-year course of depression and generalised anxiety. Br J Psychiatry 2001;**178**:18–22.
3. Sargeant JK, Bruce ML, Florio LP, Weismann MM. Factors associated with 1-year outcome of major depression in the community. Arch Gen Psychiatry 1990;**47**:519–526.
4. Keller MB. Depression: a long term illness. Br J Psychiatry 1994;**165**(Suppl. 26):9–15.
5. Ramana R, Paykel ES, Cooper Z, Hayhurst H, Saxty M, Surtees PG. Remission and relapse in major depression: a two-year prospective follow-up study. Psychol Med 1995;**25**:1161–1170.
6. Paykel ES, Cooper Z, Ramanda R, Hayhurst H. Life events, social support and marital relationships in the outcome of severe depression. Psychol Med 1996;**26**:121–133.
7. Furukawa TA, Kiturama T, Takahashi K. Time to recovery of an inception cohort with hitherto untreated unipolar depressive episodes. Br J Psychiatry 2000;**177**:331–335.
8. Nasser EH, Overholser JC. Recovery from major depression: the role of support from family, friends and spiritual beliefs. Acta Psychiatr Scand 2005;**111**:125–132.
9. Gilmer WS, Trivedi MH, Rush AJ et al. Factors associated with chronic depressive episodes: a preliminary report from the STAR-D project. Acta Psychiatr Scand 2005;**112**:425–433.
10. Spijker J, De Graaf R, Bijl RV, Beekman ATF, Ormel J, Nolen WA. Determinants of persistence of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). J Affect Disord 2004;**81**:231–240.
11. Bijl RV, Van Zessen G, Ravelli A. The Netherlands Mental Health Survey and Incidence Study (NEMESIS): objectives and design. Soc Psychiatry Psychiatr Epidemiol 1998;**33**:581–586.
12. Spijker J, Bijl RV, De Graaf R, Nolen WA. Care utilization and outcome of DSM-III-R major depression in the general population. Results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Acta Psychiatr Scand 2001;**104**:19–24.
13. Spijker J, Bijl RV, De Graaf R, Beekman ATF, Ormel H, Nolen WA. Duration of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry 2002;**181**:208–213.
14. Spijker J, Bijl RV, De Graaf R, Beekman ATF, Ormel H, Nolen WA. Functional disablity and depression. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Acta Psychiatr Scand 2004;**110**:208–214.
15. De Graaf R, Bijl RV, Smit F, Ravelli A, Vollebergh WA. Psychiatric and socio-demographic predictors of attrition in a longitudinal study: the Netherlands Mental Health

Survey and Incidence Study (NEMESIS). Am J Epidemiol 2000;**152**:1039–1047.
16. American Psychiatric Association. Diagnostic and Statistical Manual of Mental disorders (3rd edn, revised) (DSM-III-R). Washington, DC: APA, 1987.
17. Smeets RMW, Dingemans PMAJ. Composite International Diagnostic Interview (CIDI), Ver. 1.1. Amsterdam/Geneva: World Health Organization, 1993.
18. World Health Organization. Composite Diagnostic Interview (CIDI), Ver. 1.0. Geneva: World Health Organization, 1990.
19. Wittchen H-U. Reliability and validity studies of the WHO-CIDI: a critical review. J Psychiatr Res 1994;**28**:57–84.
20. Lyketsos CG, Nestadt G, Cwi J, Heithoff K, Eaton WW. The life chart interview: a standardized method to describe the course of psychopathology. Int J Methods Psychiatr Res 1994;**4**:143–155.
21. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Validity of prognostic models: when is a model clinically useful? Semin Urol Oncol 2002;**20**:96–107.
22. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;**247**:2543–2546.
23. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;**143**:29–36.
24. Efron B, Tibshirani R. An introduction to the bootstrap. Monographs on statistics and applied probability, 57. New York: Chapman & Hall, 1993.
25. Kalkman CJ, Visser K, Moen J, Bonsel GJ, Grobbee DE, Moons KG. Preoperative prediction of severe postoperative pain. Pain 2003;**105**:415–423.
26. Tiemens BG, Ormel J, Simon GE. Occurrence, recognition and outcome of psychological disorders in primary care. Am J Psychiatry 1996;**153**:636–644.
27. Ormel J, Oldehinkel AJ, Brilman E, van den Brink W. Outcome of depression and anxiety in primary care: a three-wave, 3.5-year study of psychopathology and disability. Arch Gen Psychiatry 1993;**50**:759–766.
28. Depression. Management of depression in primary and secondary care. London: Clinical guideline 23 National Institute for Clinical Excellence, 2004.
29. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;**15**:361–387.